LA-UR-21-22113
Approved for public release; distribution is unlimited.

Title: Liquid-like motions model of diffuse scattering with individual atomic B factors

Author(s): Wall, Michael E.

Intended for: Report

Issued: 2021-06-21 (rev.2)

Liquid-like motions model of diffuse scattering with individual atomic B factors

Michael E. Wall, Mail Stop B256, Los Alamos National Laboratory, Los Alamos, NM, 87505, mewall@lanl.gov

Los Alamos National Laboratory LA-UR-21-22113

March 15, 2021

Abstract
An extension of the liquid-like motions model of diffuse scattering to include individual atomic B factors is derived.

Introduction

In the liquid-like motions model (LLM) of diffuse scattering (Caspar, Clarage et al. 1988), the diffuse intensity is described using the following equation:

$$I_D(q) = e^{-|q|^2\sigma^2}|q|^2\sigma^2 I_0(q) * FT\left(e^{-\frac{|r|}{\gamma}}\right), \quad (1)$$

where q is the scattering vector, σ is the standard deviation of atom displacements along any direction, $I_0(q)$ is the squared structure factor of the unperturbed crystal, r is the separation between atom pairs, γ is a correlation length, * indicates a convolution, and FT indicates a Fourier transform. The generalization of the LLM to anisotropic displacements and correlations yields the following equation (Wall, Clarage et al. 1997):

$$I_D(q) = e^{-q\cdot V\cdot q} q\cdot V\cdot q I_0(q) * \left[\frac{8\pi \det\|G\|}{(1+|Gq|^2)^2}\right], \quad (2)$$

where V is the matrix of atomic displacement variations, and the analytic form of the Fourier transform is used, where G is the matrix of correlation lengths. A similar equation where the convolution is instead with $\frac{4\pi\det\|G\|}{1+|Gq|^2}$ (a Lorentzian) corresponds to acoustic modes, which describe streaked features in calmodulin (Wall, Clarage et al. 1997) and sharper diffuse features features near the Bragg peaks in P1 lysozyme (Meisburger, Case et al. 2020). The LLM model corresponds to the leading term in a Taylor expansion; the higher-order terms can be preserved, as was illustrated in a model similar to the LLM, but where correlations do not extend across molecular boundaries (Ayyer, Yefanov et al. 2018).

Here we extend the LLM model beyond the assumption that the displacements of all atoms are described using the same matrix V, to the case of individual atom displacements.

Model

The total squared structure factor of a crystal can be computed using the following equation:

$$I = \sum_{jk}\sum_{nm} f_m f_n^* e^{iq\cdot(R_j - R_k + x_m - x_n)} e^{iq\cdot(d_{jm} - d_{kn})}, \tag{3}$$

where R_j is the position of the origin of unit cell j, x_m is the reference position of atom m within each unit cell, and d_{jm} is the deviation of the position of atom m from the reference in unit cell j. Assuming atom displacements are statistically homogeneous, we replace the phase factor of the deviations with an average over the unit cell, then sum over k, noting that the difference between two lattice vectors is also a lattice vector:

$$I = N\sum_{j}\sum_{nm} f_m f_n^* e^{iq\cdot(r_{jmn})} \langle e^{iq\cdot(d_{jm} - d_{kn})}\rangle_k, \tag{4}$$

where

$$r_{jmn} = R_j + x_m - x_n \tag{5}$$

Now, as is standard, e.g., in calculation of the Debye-Waller factors, use the harmonic approximation to evaluate the average over the phase factor due to deviations:

$$I = N\sum_{j}\sum_{nm} f_m f_n^* e^{iq\cdot(r_{jmn})} e^{-\frac{1}{2}\langle[q\cdot(d_{jm} - d_{kn})]^2\rangle_k}. \tag{6}$$

Further evaluate the average to obtain

$$I = N\sum_{j}\sum_{nm} f_m f_n^* e^{iq\cdot(r_{jmn})} e^{-\frac{1}{2}q\cdot U_m\cdot q} e^{-\frac{1}{2}q\cdot U_n\cdot q} e^{q\cdot U_{mn}(r_{jmn})\cdot q} \tag{7}$$

The exponential factors correspond to Debye-Waller factors for individual atoms (involving U_m and U_n) and a factor for the correlated displacements of two atoms (involving U_{mn}). We will make some assumptions about the U_{mn} factor to derive the modified LLM model, but first, make sure the incoherent scattering ($j = 0$ and $m = n$) is properly accounted for when choosing an arbitrary U_{mn} that might not precisely match the individual atomic displacements:

$$\begin{aligned} I = {} & N\sum_{j}\sum_{nm} f_m f_n^* e^{iq\cdot(r_{jmn})} e^{-\frac{1}{2}q\cdot U_m\cdot q} e^{-\frac{1}{2}q\cdot U_n\cdot q} e^{q\cdot U_{mn}(r_{jmn})\cdot q} \\ & + N\sum_{m} |f_m|^2 \left(1 - e^{-q\cdot U_m\cdot q} e^{q\cdot U_{mm}(0)\cdot q}\right) \end{aligned} \tag{8}$$

Now separate the Bragg term and isolate the diffuse intensity from both the incoherent correction and Bragg:

$$I = N\sum_j\sum_{nm} f_m f_n^* e^{iq\cdot(r_{jmn})} e^{-\frac{1}{2}q\cdot U_m\cdot q} e^{-\frac{1}{2}q\cdot U_n\cdot q} + N\sum_j\sum_{nm} f_m f_n^* e^{iq\cdot(r_{jmn})} e^{-\frac{1}{2}q\cdot U_m\cdot q} e^{-\frac{1}{2}q\cdot U_n\cdot q}\left(e^{q\cdot U_{mn}(r_{jmn})\cdot q} - 1\right) + N\sum_m |f_m|^2\left(1 - e^{-q\cdot U_m\cdot q} e^{q\cdot U_{mm}(0)\cdot q}\right) \tag{9}$$

From which we identify the first term as corresponding to the Bragg:

$$I_B = N\sum_j\sum_{nm} f_m f_n^* e^{iq\cdot(r_{jmn})} e^{-\frac{1}{2}q\cdot U_m\cdot q} e^{-\frac{1}{2}q\cdot U_n\cdot q}, \tag{10}$$

the middle term corresponding to the diffuse:

$$I_D = N\sum_j\sum_{nm} f_m f_n^* e^{iq\cdot(r_{jmn})} e^{-\frac{1}{2}q\cdot U_m\cdot q} e^{-\frac{1}{2}q\cdot U_n\cdot q}\left(e^{q\cdot U_{mn}(r_{jmn})\cdot q} - 1\right), \tag{11}$$

and the last term corresponding to the incoherent correction:

$$I_{inc} = N\sum_m |f_m|^2\left(1 - e^{-q\cdot U_m\cdot q} e^{q\cdot U_{mm}(0)\cdot q}\right). \tag{12}$$

Now, the key assumption of the LLM is that the $U_{mn}(r_{jmn})$ has the same form for all pairs of atoms, and only depends on the separation vector r_{jmn}. In this case $U_{mn}(r_{jmn})$ can be written as

$$U_{mn}(r_{jmn}) = Vc(r_{jmn}), \tag{13}$$

where V is a (potentially anisotropic) matrix of variations and $c(r_{jmn})$ is a scalar function with values in the range [-1,1] describing the correlation vs separation. Making this substitution and using a Taylor expansion for the exponential yields

$$N\sum_{l=1}^{\infty}\frac{(q\cdot V\cdot q)^l}{l!}\sum_j\sum_{nm} e^{iq\cdot(r_{jmn})} f_m f_n^* e^{-\frac{1}{2}q\cdot U_m\cdot q} e^{-\frac{1}{2}q\cdot U_n\cdot q} c^l(r_{jmn}), \tag{14}$$

which, using the Dirac delta function, is equivalent to

$$\sum_{l=1}^{\infty}\frac{(q\cdot V\cdot q)^l}{l!}\int d^3r\, e^{iq\cdot r} c^l(r) N\sum_j\sum_{nm}\delta(r - r_{jmn}) f_m f_n^* e^{-\frac{1}{2}q\cdot U_m\cdot q} e^{-\frac{1}{2}q\cdot U_n\cdot q}. \tag{15}$$

Note that

$$\int d^3r\, e^{iq\cdot r} N \sum_j \sum_{nm} \delta\big(r - r_{jmn}\big) f_m f_n^* e^{-\frac{1}{2}q\cdot U_m\cdot q} e^{-\frac{1}{2}q\cdot U_n\cdot q} \tag{16}$$
$$= N \sum_j \sum_{nm} e^{iq\cdot(r_{jmn})} f_m f_n^* e^{-\frac{1}{2}q\cdot U_m\cdot q} e^{-\frac{1}{2}q\cdot U_n\cdot q} = I_B,$$

which identifies the double sum in Eq. (15) as the Patterson of the mean electron density of the crystal, $P(r)$. We therefore have

$$I_D = \sum_{l=1}^{\infty} \frac{(q\cdot V\cdot q)^l}{l!} \int d^3r\, e^{iq\cdot r} c^l(r) P(r), \tag{17}$$

or, by the convolution theorem,

$$I_D = \sum_{l=1}^{\infty} \frac{(q\cdot V\cdot q)^l}{l!} I_B * \int d^3r\, e^{iq\cdot r} c^l(r). \tag{18}$$

which, to first order, is

$$I_D = q\cdot V\cdot q I_B * FT[c(r)]. \tag{19}$$

Compared to Eq. (2), this equation is missing the leading factor $e^{-q\cdot V\cdot q}$. Eq. (2) otherwise has the same form as the first order term in this expansion, but with the unperturbed squared structure factor I_0 (computed without individual atom B factors) replaced by the Bragg intensity I_B (computed using the individual atom B factors).

As an example, the following is the equation for the isotropic LLM using individual B factors, in which the correlation decreases exponentially with distance:

$$I_D = |q|^2\sigma^2 I_B * FT e^{-\frac{|r|}{\gamma}}, \tag{20}$$

yielding

$$I_D(q) = |q|^2\sigma^2 I_B(q) * \frac{8\pi\gamma^3}{[1+(\gamma|q|)^2]^2}. \tag{21}$$

Similar to the original LLM, in this individual B factor LLM, σ corresponds to the portion of the atom displacements that obey an exponential falloff of the correlations with distance. Note that, to first order, changes in σ simply scale the entire diffuse pattern and do not change the shape of the distribution. Obtaining an estimate of σ requires placing the diffuse intensity on an absolute scale. Methods for placing the diffuse intensity on an absolute scale were recently developed in a study of P1 lysozyme (Meisburger, Case et al. 2020).

Conclusion

Individual atomic B factors can be included in the LLM model of diffuse scattering by eliminating the overall Debye-Waller factor $\left(e^{-|q|^2\sigma^2}\right)$, and replacing $I_0(q)$, the squared structure factor of the unperturbed crystal, with $I_B(q)$, the Bragg intensity computed from the participating atoms.

Acknowledgment

I am grateful to Zhen Su and Chuck Yoon for proofreading this document, and for Zhen Su, Chuck Yoon, and Mark Wilson for discussions.

References

Ayyer, K., O. M. Yefanov and H. N. Chapman (2018). Structure Determination by Continuous Diffraction from Imperfect Crystals. X-ray Free Electron Lasers: A Revolution in Structural Biology. S. Boutet, P. Fromme and M. S. Hunter. Cham, Springer International Publishing**:** 253-300.

Caspar, D. L., J. Clarage, D. M. Salunke and M. Clarage (1988). "Liquid-like movements in crystalline insulin." Nature **332**(6165): 659-662.

Meisburger, S. P., D. A. Case and N. Ando (2020). "Diffuse X-ray scattering from correlated motions in a protein crystal." Nat Commun **11**(1): 1271.

Wall, M. E., J. B. Clarage and G. N. Phillips (1997). "Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering." Structure **5**(12): 1599-1612.